

Ontologías y Semántica en el Proceso de Visualización

Martín L. Larrea, Sebastián Escarza, Dana K. Urribarri y Silvia M. Castro
Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica (VyGLab)
Departamento de Ciencias e Ingeniería de la Computación
Universidad Nacional del Sur
Bahía Blanca, Argentina
{mll, se, dku, smc}@cs.uns.edu.ar

Resumen

La Visualización como área se encuentra en un estado en el cual existe un consenso parcial acerca de sus principales conceptos. Sin embargo, aún no se ha alcanzado un estado de madurez suficiente en el área y el establecimiento de un vocabulario formal común sigue siendo un objetivo pendiente.

Esta línea de investigación tiene como objetivo crear un modelo de visualización que considere la semántica de los datos, del contexto y de las etapas del “Modelo Unificado de Visualización” (MUV) para poder asistir al usuario en el seteo de los parámetros de la visualización logrando así una mejor representación visual. Este nuevo modelo, que se presenta como una extensión del MUV, deberá formalizar el modelo de referencia mediante una Ontología de Visualización, incorporar una Ontología de Datos e incorporar una capa de inferencia que le permita derivar nueva información, a partir de la ya conocida.

Palabras clave: *Ontología, Representación Formal, Visualización basada en semántica, Visualización.*

Contexto

Este trabajo continúa la línea presentada en [LEUC12] y se lleva a cabo en el Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica (VyGLab) del Departamento de Ciencias e Ingeniería de la Computación de la Universidad Nacional del Sur.

La línea de Investigación presentada está inserta en el proyecto y “Representaciones Visuales e Interacciones para el Análisis Visual de Grandes Conjuntos de Datos”

(24/N020), dirigido por la Doctora Silvia Castro y , financiado por la Secretaría General de Ciencia y Tecnología de la Universidad Nacional del Sur; y acreditado por la Universidad Nacional del Sur, Bahía Blanca.

Introducción

Actualmente existe un gran número de modelos de referencia en Visualización [CM97, Chi00, Shn04, BN07] que, de diversas maneras, han identificado y delineado los principales componentes y procesos que sufren los datos para ser visualizados. En particular, en el seno del VyGLab hemos estado desarrollando el Modelo Unificado de Visualización (MUV) [MCFE03], un modelo que constituye un marco conceptual de referencia en términos del cual ubicar los procesos y estados intermedios de los datos y definir las interacciones explícitamente.

Si bien estos esfuerzos por compilar, caracterizar y clasificar los aspectos más relevantes del área han sido y siguen siendo fructíferos, aún no existe un consenso definitivo que permita consolidar una teoría de base en Visualización.

La incorporación de una semántica bien definida en el proceso de Visualización ya ha sido identificada como una necesidad [DBD*05, VO04, CEH*09] en el sentido que permitiría formalizar dicho proceso, estableciendo un vocabulario común que permita a los usuarios plantear sus requerimientos, y a los diseñadores de herramientas de visualización, expresar las transformaciones de los datos desde su obtención hasta la construcción de la vista, las operaciones que deben proveerse y las

formas de interacción posibles entre los usuarios y la visualización

Para lograr una formalización consensuada del proceso de Visualización, es preciso describir axiomáticamente el marco de referencia provisto por los modelos mencionados anteriormente. En este contexto, las ontologías surgen como una herramienta natural para estos fines.

Un ejemplo de los beneficios de la formalización del proceso de visualización se encuentra en la formalización de los tipos de datos, es decir una Ontología de Datos. Contar con una ontología que clasifique los conjuntos de datos a visualizar es una guía que asiste a la hora de elegir la técnica de visualización apropiada para determinado conjunto de datos. En la literatura es posible encontrar taxonomías, una versión menos formal de la ontología, de datos pero no son presentadas desde un punto de vista de visualización.

Por otro lado, el desafío de una visualización es encontrar una metáfora visual que permita entender y percibir en forma efectiva un conjunto de datos. Para este propósito es útil contar con cuantificaciones de diferentes aspectos de cada técnica de visualización. Una visualización debe proveer también un conjunto de interacciones a partir de las cuales el usuario explorará el conjunto de datos con una mínima carga cognitiva. La tecnología computacional actual permite la exploración de grandes conjuntos de información. Por un lado, esta situación es extremadamente útil pero la creciente cantidad de información genera una sobrecarga cognitiva. Mientras que el poder computacional ha crecido en forma exponencial, la habilidad para interactuar con dichos datos solo se ha incrementado en forma lineal.

Hoy en día, una gran variedad de usuarios acceden, extraen y muestran información que está distribuida sobre diferentes puntos, con diferentes tipos, formas y contenidos. En muchos casos, el usuario debe tener un control activo sobre el proceso de visualización pero, aún en

este caso, es difícil obtener una visualización efectiva. Es común que la información que se desea representar no tenga una manifestación visual obvia, ante esta situación el proceso de mapeo del conjunto de datos a la vista puede llegar a ser no trivial [CMT01].

Líneas de Investigación y Desarrollo

Este trabajo presenta tres ejes de investigación las cuales se entrelazan y tienen varios puntos en común. Tales ejes son:

1. “Formalización del proceso de Visualización mediante el uso de Ontologías”.
2. “Caracterización de Conjuntos de Datos”.
3. “Caracterización de Técnicas de Visualización”.
4. “Visualización Basada en Semántica”.

Formalización del Proceso de Visualización mediante el uso de Ontologías

Este eje es una continuación directa de la línea de investigación presentada en ediciones anteriores de este Workshop [ECM10], en la cual se plantea la necesidad de definir una Ontología de Visualización y de brindar soporte a dicha especificación formal a través de una plataforma de software que facilite la integración de los diversos componentes.

Este trabajo integra el uso de ontologías en Visualización. Las ontologías, si bien se originaron hace tiempo en el área de Representación del Conocimiento, han cobrado nueva relevancia con el surgimiento de proyectos como la Web Semántica. A raíz de ello, numerosos estándares han sido definidos. Especificaciones como RDF, RDF Schema y OWL (el Lenguaje de Ontologías de la Web) permiten definir vocabularios estructurados

que modelan jerarquías de herencia entre conceptos, relaciones, restricciones y reglas, que permiten un tratamiento más riguroso del conocimiento al poseer una semántica formal asociada.

En cuanto a la integración de semántica en visualización, aún no se cuenta con abordajes sistemáticos al problema y se identifican dos enfoques principales. Por un lado, se utilizan representaciones formales para mejorar la integración, consulta y descripción de los datos del usuario como una forma de enriquecer la visualización, y por el otro, trabajos más relacionados con el que proponemos utilizan descripciones semánticas, aunque en principio algo limitadas, del proceso de visualización en sí mismo para ayudar al usuario en la definición y configuración del mismo y en la selección de una técnica de visualización apropiada.

Caracterización de Conjuntos de Datos

Este eje se centra en el desarrollo de una clasificación de los conjuntos de datos orientada a la visualización. Esta clasificación debe brindar suficiente información sobre cuáles son las características con las que debe contar la técnica de visualización que se emplee para visualizar cada categoría de datos.

El desafío es encontrar métricas que, no sólo permitan evaluar en forma lo más sencilla posible cada uno de los aspectos importantes a tener cuenta, sino que también permitan una clasificación conveniente de los datos.

Caracterización de Técnicas de Visualización

Si bien a la hora de visualizar conjuntos de datos pequeños la mayoría de las técnicas resultan adecuadas, la situación cambia cuando se quieren visualizar grandes conjuntos de datos: una mala elección de la técnica a utilizar o de sus atributos y el resultado obtenido puede no

ser satisfactorio. Una posible solución a este problema es contar con alguna herramienta que permita predecir qué técnica es más "conveniente" usar para obtener información de un determinado conjunto de datos.

En este contexto se buscan desarrollar métricas particulares para cada técnica de visualización con el objetivo de caracterizar numéricamente algún aspecto del resultado final de la aplicación de dicha técnica sobre un determinado conjunto de datos.

Visualización Basada en Semántica

Debido a que el objetivo de una visualización es lograr una representación que ayude al usuario a interpretar un conjunto de datos y comunicar su significado, es importante controlar el mapeo de las dimensiones físicas a las preceptuales; un usuario inexperto podría utilizar un mapeo incorrecto afectando negativamente la visualización resultante. Una estrategia para mejorar esta situación es guiar al usuario en la selección de los diferentes parámetros involucrados en la visualización a través de información semántica.

Resultados y Objetivos

A nivel de la primera línea de investigación presentada, la transformación de los objetivos en resultados sigue las etapas delineadas en [ECM10].

Al momento de esta publicación se cuenta con una definición parcial de la Ontología de Visualización. Concretamente, la arquitectura general de la ontología ya ha sido definida así como los principales elementos de la misma.

Adicionalmente, también se ha definido un conjunto de conceptos destinados a caracterizar el espacio de datos del usuario y la representación visual, y se continúa trabajando en la definición de los conceptos ontológicos necesarios para describir el proceso de visualización como una red de etapas interconectadas, así como en las

definiciones necesarias para especificar las interacciones entre el usuario y la visualización.

Sobre la segunda línea presentada, hasta el momento se han evaluado cuales son las posibles métricas que se empelarán para categorizar los datos. En este sentido, se seleccionaron varias métricas globales sobre grafos, métricas sobre tablas de información y diferentes medidas de dispersión.

La tercera línea se desprende del trabajo realizado en la línea anterior y persigue el desarrollo de métricas específicas para cada técnica de visualización. Hasta el momento se han analizado métricas particulares para Scatterplots.

Finalmente, sobre la tercera línea de investigación, se ha logrado incluir la semántica de los datos en el mapeo de datos a colores en una representación visual [LMC10]. Gracias a este trabajo, esta etapa del proceso de visualización ya no requiere de la participación del usuario; las decisiones de qué color usar para cada dato se determinan a partir de un razonador semántico.

Formación de Recursos Humanos

En lo concerniente a la formación de recursos humanos se detallan las tesis concluidas y en desarrollo relacionadas con la línea de investigación presentada:

Tesis Concluidas

- Tesis Doctoral. Martín L. Larrea. Tema: *Visualización Basada en Semántica*. Dirección: Dra. Silvia Castro.

Tesis en Desarrollo

- Tesis Doctoral. Sebastián Escarza. Tema: *Ontologías de Visualización*. Dirección: Dra. Silvia Castro.

- Tesis Doctoral. Dana Urribarri. Tema: *Escalabilidad Visual*. Dirección: Dra. Silvia Castro.
- Tesis Doctoral. María Luján Ganuza. Tema: *Servicios Web en Visualización de Información*. Dirección: Dra. Silvia Castro.

Referencias

[BN07] K. Brodlie and N. M. Noor. Visualization Notations, Models and Taxonomies. pages 207–212, Bangor, United Kingdom, 2007. Eurographics Association.

[CEH*09] M. Chen, D. Ebert, H. Hagen, R. S. Laramée, R. van Liere, K.-L. Ma, W. Ribarsky, G. Scheuermann, and D. Silver. Data, information, and knowledge in visualization. *IEEE Comput. Graph. Appl.*, 29(1):12–19, 2009.

[CM97] S. K. Card and J. Mackinlay. The structure of the information visualization design space. In *INFOVIS '97: Proceedings of the 1997 IEEE Symposium on Information Visualization (InfoVis '97)*, page 92, Washington, DC, USA, 1997. IEEE Computer Society.

[CMT01] Carpendale, M. S. T. 2001. Considering Visual Variables as a Basis for Information Visualization. Technical Report. University of Calgary, Department of Computer Science.

[Chi00] E. H. Chi. A taxonomy of visualization techniques using the data state reference model. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis'00)*, page 6975. IEEE Computer Society Press, 2000.

[DBD*05] D. J. Duke, K. W. Brodlie, D. A. Duce, and I. Herman. Do you see what I mean? *IEEE. Computer Graphics and Applications*, 25(3):6–9, 2005.

[ECM07] S. Escarza, S. Castro, S. Martig. Ontologías de Visualización. IX Workshop de Investigadores en Ciencias de la Computación (WICC2007) pp. 275-278. Universidad Nacional de la Patagonia San Juan Bosco. 3 y 4 de Mayo de 2007. Trelew. Chubut. Argentina. ISBN 978-950-763-075-0.

[ECM08] S. Escarza, S. Castro, S. Martig. Desafíos en el camino desde el Modelo Unificado de Visualización hasta la construcción de visualizaciones. X Workshop de Investigadores en Ciencias de la Computación (WICC2008) pp. 297-300. Universidad Nacional de la Pampa. 5 y 6 de Mayo de 2008. Gral. Pico. La Pampa. Argentina. ISBN 978-950-863-101-5.

[ECM10] S. Escarza, S. Castro, S. Martig. Formalización del Proceso de Visualización mediante el uso de Ontologías. XII Workshop de Investigadores en Ciencias de la Computación (WICC 2010) pp. 261-265. Universidad Nacional de la Patagonia San Juan Bosco. 5 y 6 de mayo de 2010. El Calafate. Santa Cruz. Argentina. ISBN: 978-950-34-0652-6.

[LEUC12] Martín L. Larrea, Sebastián Escarza, Dana K. Urribarri, Sergio R. Martig y Silvia M. Castro. Ontologías y Semántica en el Proceso de Visualización. XIV Workshop de Investigadores en Ciencias de la Computación (WICC 2012), pp. 302–306. Universidad Nacional de Misiones, Posadas, Misiones (abril 2012). ISBN 978-950-766-082-5

[LMC10] M. Larrea, S. Martig and S. Castro. Semantics-based Color Assignment in Visualization. Journal of Computer Science & Technology. Vol. 10 - No. 1 – April 2010 - ISSN 1666-6038.

[LMC10*] M. Larrea, S. Martig and S. Castro. Formalización del Proceso de Visualización Basada en Semántica. XII Workshop de Investigadores en Ciencias de

la Computación (WICC 2010) pp. 270-274. Universidad Nacional de la Patagonia San Juan Bosco. 5 y 6 de mayo de 2010. El Calafate. Santa Cruz. Argentina. ISBN: 978-950-34-0652-6.

[MCFE03] Martig S., Castro S., Fillottrani P., Estévez E., Un Modelo Unificado de Visualización, Proceedings, pp. 881-892, 9º Congreso Argentino de Ciencias de la Computación. 6 al 10 de Octubre de 2003. La Plata. Argentina.

[Shn04] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. IEEE Symposium on Visual Languages, 0:336, 1996.

[UCM10] D. K. Urribarri, S. Castro, S. Martig. Caracterización de Conjuntos de Datos. XII Workshop de Investigadores en Ciencias de la Computación (WICC 2010) pp. 266-269. Universidad Nacional de la Patagonia San Juan Bosco. 5 y 6 de mayo de 2010. El Calafate. Santa Cruz. Argentina. ISBN: 978-950-34-0652-6.

[VO04] Reporte del Visualisation Ontology Workshop realizado en Nacional e-Science Centre el 7 y 8 de Abril de 2004. <http://www.nesc.ac.uk/esi/events/393/>